



A comparison of the performance of 4D-Var in an explicit and implicit version of a nonlinear barotropic ocean model

A. D. Terwisscha van Scheltinga, H. A. Dijkstra

► To cite this version:

A. D. Terwisscha van Scheltinga, H. A. Dijkstra. A comparison of the performance of 4D-Var in an explicit and implicit version of a nonlinear barotropic ocean model. *Nonlinear Processes in Geophysics*, 2007, 14 (6), pp.763-776. hal-00302938

HAL Id: hal-00302938

<https://hal.science/hal-00302938>

Submitted on 30 Nov 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A comparison of the performance of 4D-Var in an explicit and implicit version of a nonlinear barotropic ocean model

A. D. Terwisscha van Scheltinga and H. A. Dijkstra

Institute for Marine and Atmospheric research Utrecht, Department of Physics and Astronomy, Utrecht University, Princetonplein 5, 3584 CC, Utrecht, The Netherlands

Received: 4 July 2007 – Revised: 1 November 2007 – Accepted: 3 November 2007 – Published: 30 November 2007

Abstract. A comparison is made of the performance of the four-dimensional variational data assimilation (4D-Var) method in an explicit and implicit version of a barotropic quasi-geostrophic model of the wind-driven double-gyre ocean circulation. As is well known, implicit methods have the advantage that relatively large time steps can be taken with respect to explicit methods, but the computational costs of each time step is larger. We focus here on two issues: (i) the computational efficiency in the range of time steps where the chosen explicit method is still numerically stable and (ii) the performance of 4D-Var in the implicit model for time steps out of reach for the explicit model. For the same time step Δt and the same number of points n per assimilation interval, the analyses in the implicit model is always more accurate than that in the explicit model. Due to this property the use of 4D-Var combined with the implicit model can be computationally more efficient than its use in the explicit model.

1 Introduction

The four-dimensional variational data assimilation method, 4D-Var, is now widely applied in meteorology and physical oceanography. It is a method in which information that is present in observations is combined with the evolution determined by a particular ocean, atmosphere or climate model. The 4D-Var method is routinely applied at ECMWF in weather forecasting (Rabier et al., 2000; Mahfouf and Rabier, 2000; Klinker et al., 2000). In operational oceanography, for example within the French Mercator project (Weaver et al., 2003; Vialard et al., 2003), the use of observations to initialize ocean circulation models results in better forecasts. The Estimating the Circulation and Climate of the

Ocean (ECCO) consortium applied 4D-Var to estimate the ocean state using WOCE data and NCEP reanalysis of the surface fluxes (Stammer et al., 2000, 2002a,b, 2003). The LSG model (Hamburg Large Scale Geostrophic Model) was used together with 4D-Var by Wenzel et al. (2001) to study the annual cycle of the global ocean circulation. The 4D-Var method was also used for paleonutrient data analysis of the glacial Atlantic (Winguth et al., 2000). In the 4D-Var approach, a cost function is minimized by varying the initial condition or the forcing of the model. Parameters controlling the solution can also be varied, for example diffusivities (Stammer, 2005) or eddy stresses (Ferreira et al., 2005). The cost function measures the distance between the data and a state vector at a sequence of times. The so-called analysis is that state which minimizes the cost function and the minimization procedure requires the evaluation of the gradient of the cost function. In general, this gradient is calculated by using both a forward and an adjoint model. Although there are now compilers which generate actual computer code of an adjoint model given the code of the forward model (Giering and Kaminski, 1998), the formulation of an adjoint model is in most cases a nontrivial and time-consuming process. With the increase in the length of data sets, data-assimilation becomes feasible to better simulate phenomena with a longer time scale, such as interannual variations of the Gulf Stream path or even decadal variability in the North Atlantic. For these type of computations, the implementation of 4D-Var methods in explicit models have the drawback that the time step is restricted by the explicit time stepping scheme. Even if one uses monthly averaged data, the time step of the ocean model has to be small (smaller than an hour at reasonable resolution) to carry out the assimilation. Such long term data assimilation problems motivate to search for alternative 4D-Var implementations in which relatively large time steps can be taken. By tackling problems associated with the stability and successive bifurcations of large-scale ocean flows, fully-implicit ocean models have been developed over

Correspondence to: A. D. Terwisscha van Scheltinga
(a.d.terwisschavanscheltinga@phys.uu.nl)

the last decade. For example, implicit quasi-geostrophic and shallow-water models of the wind-driven ocean circulation have been used to investigate the bifurcation behavior of the double-gyre circulation (Dijkstra and Katsman, 1997; Schmeits and Dijkstra, 2000). A hierarchy of fully-implicit models of the thermohaline ocean circulation has helped clarify the role of different equilibria in the hysteresis behavior of the global ocean circulation (Dijkstra et al., 2004). The immediate advantage of these methods is that much larger time steps can be taken than with explicit methods. A few years ago 4D-Var was implemented in fully implicit models (Terwisscha van Scheltinga and Dijkstra, 2005). In this implementation, the adjoint model is easily derived from the implicit time-stepping scheme, and the choice of the time step is not limited by numerical stability but by accuracy. In this paper, we compare the performance of the implementation of 4D-Var in an implicit version (abbreviated below with *i*4D-Var) of a barotropic quasi-geostrophic model of the double-gyre wind-driven circulation with 4D-Var applied to an explicit version (abbreviated below with *e*4D-Var) of the same model. The aim of the comparison is to investigate whether implicit methods provide useful alternatives in problems where variational data-assimilation techniques are used.

2 Model and methods

In the first subsection below (Sect. 2.1), we provide the model of the wind-driven ocean circulation which is used in this study. Next, we provide a basic overview of the 4D-Var method (Sect. 2.2) such that the differences between the implementation of 4D-Var in the implicit and explicit versions of the ocean model in Sect. 2.3 can be explained more easily.

2.1 Barotropic wind-driven ocean flows

Consider a rectangular ocean basin of size $L \times L$ having a constant depth D . The basin is situated on a midlatitude β -plane with a central latitude $\theta_0 = 45^\circ$ N and Coriolis parameter $f_0 = 2\Omega \sin \theta_0$, where Ω is the rotation rate of the Earth. The meridional variation of the Coriolis parameter at the latitude θ_0 is indicated by β_0 . The density ρ of the water is constant and the flow is forced at the surface through a wind-stress vector $\mathbf{T} = \tau_0[\tau^x(x, y), \tau^y(x, y)]$. The governing equations are non-dimensionalized using a horizontal length scale L , a vertical length scale D , a horizontal velocity scale U , the advective time scale L/U and a characteristic amplitude of the wind-stress vector, τ_0 . The effect of deformations of the ocean-atmosphere interface on the flow is neglected.

The dimensionless barotropic quasi-geostrophic model of the flow for the vorticity ζ and the geostrophic streamfunction ψ is (Pedlosky, 1987)

$$\left[\frac{\partial}{\partial t} + u \frac{\partial}{\partial x} + v \frac{\partial}{\partial y} \right] [\zeta + \beta y] = Re^{-1} \nabla^2 \zeta + \alpha_\tau \left(\frac{\partial \tau^y}{\partial x} - \frac{\partial \tau^x}{\partial y} \right), \quad (1a)$$

$$\zeta = \nabla^2 \psi, \quad (1b)$$

where the dimensionless horizontal velocities are given by $u = -\partial \psi / \partial y$ and $v = \partial \psi / \partial x$. The parameters in Eq. (1a) are the Reynolds number Re , the planetary vorticity gradient parameter β and the wind-stress forcing strength α_τ . These parameters are defined as:

$$Re = \frac{UL}{A_H}; \quad \beta = \frac{\beta_0 L^2}{U}; \quad \alpha_\tau = \frac{\tau_0 L}{\rho D U^2} \quad (2)$$

where g is the gravitational acceleration and A_H is the lateral friction coefficient. We assume no-slip conditions on the east-west boundaries and slip on the north-south boundaries. The boundary conditions are therefore given by

$$x = 0, x = 1 : \psi = \frac{\partial \psi}{\partial x} = 0, \quad (3a)$$

$$y = 0, y = 1 : \psi = \zeta = 0. \quad (3b)$$

The wind-stress forcing is prescribed as

$$\tau^x(x, y) = \frac{-1}{2\pi} \cos 2\pi y, \quad (4b)$$

$$\tau^y(x, y) = 0 \quad (4b)$$

and the zonal wind stress is symmetric with respect to the mid-axis of the basin (the standard double-gyre case). When the horizontal velocity scale is based on a Sverdrup balance of the flow, i.e.,

$$U = \frac{\tau_0}{\rho D \beta_0 L}, \quad (5)$$

it follows that $\alpha_\tau = \beta$ and two free parameters result (Pedlosky, 1987), for example the dimensionless boundary layer thicknesses $\delta_I^2 = 1/\beta$ and $\delta_M^3 = 1/(\beta Re)$. A standard set of parameter values has been chosen (Table 1) that are similar to those in Dijkstra and Katsman (1997) and for these parameters, $\alpha_\tau = \beta = 2.8 \times 10^3$.

For the parameters as in Table 1, Dijkstra and Katsman (1997) showed that different flow regimes exist when Re is varied. For $Re < 30$, the quasi-geostrophic model has one unique stable steady state. The streamfunction ψ of this steady state is anti-symmetric with respect to the mid-axis of the basin and a solution for $Re = 20$ is shown in Fig. 1a. Two asymmetric stable steady-state solutions, one with a downward jet-displacement (the jet-down solution shown in Fig. 1b) and the jet-up solution (Fig. 1c), exist

Table 1. Standard values of the parameters for the barotropic quasi-geostrophic ocean model.

Parameter	Value
L	1.0×10^6 m
U	7.1×10^{-3} ms ⁻¹
D	7.0×10^2 m
β_0	2.010^{-11} (ms) ⁻¹
f_0	1.0×10^{-4} s ⁻¹
g	9.8 ms ⁻²
ρ	1.0×10^3 kgm ⁻³
τ_0	1.0×10^{-1} Pa

for $30 < Re < 52$. Near $Re=52$ both asymmetric states become unstable due to the occurrence of Hopf bifurcations; for $52 < Re < 74$ stable periodic orbits exist. The solutions become quasi-periodic for $Re > 74$ and irregular for higher values of Re ; the route to chaos is through a homoclinic orbit (Simonnet et al., 2005).

2.2 The 4D-Var method

The incremental formulation of the 4D-Var method as described in Courtier et al. (1994) is used and the notation is adapted from Ide et al. (1997). Let \mathbf{w} be the state vector consisting of model variables that are to be estimated by combining model dynamics and observations. If \mathbf{w}^b is the background state and $\delta\mathbf{w}$ is the increment on the background state, then we want to determine $\delta\mathbf{w}$ such that the resulting state \mathbf{w} defined by

$$\mathbf{w} = \mathbf{w}^b + \delta\mathbf{w} \quad (6)$$

is “close” to observations. In the 4D-Var approach, the analysis \mathbf{w}^a is defined as the state vector which minimizes both the distance to the background $\mathbf{w}^b(t_0)$ and to the time-sequence of observations $\mathbf{y}_i, i=1, \dots, n$ in the interval $t_0 \leq t_i \leq t_n$. Hence, this defines a cost function J as (Courtier et al., 1994):

$$J(\delta\mathbf{w}) = \delta\mathbf{w}^T \mathbf{B}^{-1} \delta\mathbf{w} + \sum_{i=0}^n \mathbf{d}_i^T \mathbf{R}_i^{-1} \mathbf{d}_i, \quad (7)$$

where the departures \mathbf{d}_i are defined as:

$$\mathbf{d}_i = \mathbf{y}_i - \mathbf{H}_i \mathbf{M}(t_i, t_0)(\mathbf{w}^b(t_0)) - \mathbf{H}_i \mathbf{M}(t_i, t_0) \delta\mathbf{w}(t_0). \quad (8)$$

The matrices \mathbf{B} and \mathbf{R}_i in Eq. (7) are the covariances of the background and the observation errors. The operator $\mathbf{M}(t_i, t_0)$ in Eq. (8) represents the evolution operator, such that

$$\mathbf{w}(t_i) = \mathbf{M}(t_i, t_0)(\mathbf{w}(t_0)), \quad (9)$$

and \mathbf{H}_i in Eq. (8) is the observation operator. The linearization of the operators $\mathbf{M}(t_i, t_0)$ and \mathbf{H}_i around the background

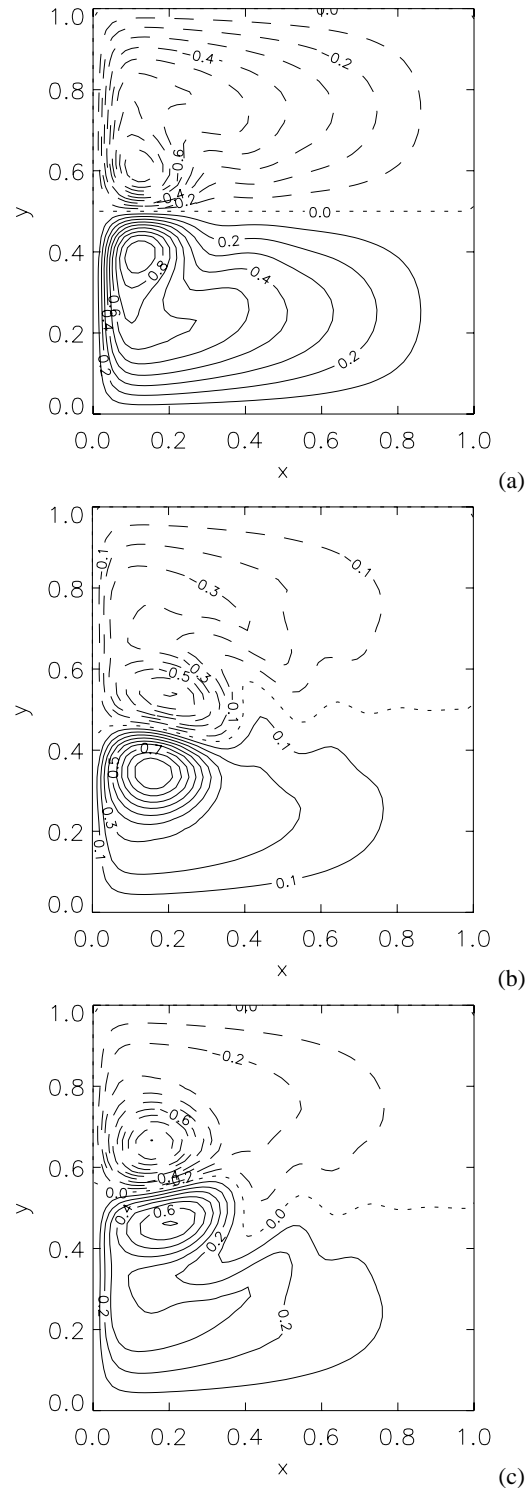


Fig. 1. Streamfunction ψ of (a) the anti-symmetric steady state for $Re=20$, (b) the jet-down steady state for $Re=50$ and (c) the jet-up steady state for $Re=50$. The contour values are scaled with respect to a maximum of $\psi=2.2$ for (a), which represents a transport of 5.5 Sv; and a maximum of $\psi=1.1$ for (b, c), which represents a transport of 10.9 Sv. The contour interval is 0.2.

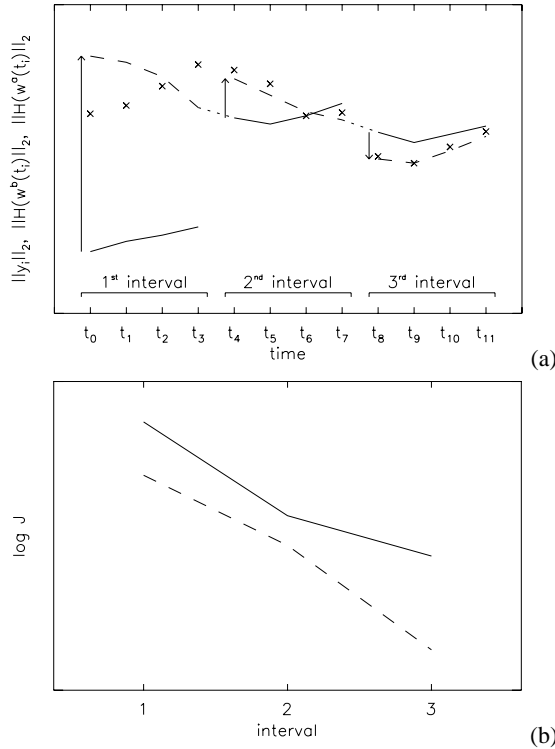


Fig. 2. Sketch of the 4D-Var method, where an assimilation interval has been divided into three subintervals. (a) $\|H(\mathbf{w}^b(t_i))\|_2$ and $\|H(\mathbf{w}^a(t_i))\|_2$, the L_2 -norm of the projection of the background $\mathbf{w}^b(t_i)$ (solid) and analysis $\mathbf{w}^a(t_i)$ (dashes) on the observations space; the L_2 -norm of the observations $\|\mathbf{y}_i\|_2$ (crosses) and the optimal increments $\delta \mathbf{w}^a$ (arrows). (b) the initial (solid) and final (dashed) value of the cost function.

state are denoted by $\mathbf{M}(t_i, t_0)$ and \mathbf{H}_i , respectively. If $\delta \mathbf{w}^a$ is defined as the solution of the optimization problem

$$\delta \mathbf{w}^a = \min_{\delta \mathbf{w}} J(\delta \mathbf{w}), \quad (10)$$

then the analysis is given by

$$\mathbf{w}^a(t_0) = \mathbf{w}^b(t_0) + \delta \mathbf{w}^a. \quad (11)$$

To solve the optimization problem Eq. (10) on each subinterval, the gradient of the cost function J in Eq. (7), i.e.,

$$\nabla J(\delta \mathbf{w}) = 2\mathbf{w}^{-1}\delta \mathbf{w} - \sum_{i=0}^n \mathbf{M}^T(t_i, t_0) \mathbf{H}_i^T \mathbf{R}_i^{-1} \mathbf{d}_i, \quad (12)$$

has to be calculated.

To clarify the terminology used below, an illustration of the 4D-Var method has been provided in Fig. 2a. In this figure, the observations (crosses) are shown on an example assimilation interval $t_1 \leq t_i \leq t_{12}$. This interval has been divided into three subintervals, each with four points ($n=4$). For every interval the background trajectory $\mathbf{w}^b(t_i)$

(solid), the optimal increment $\delta \mathbf{w}^a$ (arrows) and the analysis $\mathbf{w}^a(t_i)$ (dashed) are shown. The background on the first interval is given. For the other intervals, the background is calculated from the analysis on the previous interval: $\mathbf{w}^b(t_{n+1}) = \mathbf{M}(t_{n+1}, t_n)(\mathbf{w}^a(t_n))$. On each interval the minimization problem Eq. (10) is solved. Due to the dependence of the cost function on the background, the increment and the observations, the initial and the final value of the cost function will vary over the subintervals (Fig. 2b).

2.3 Explicit and implicit implementations

The equations Eq. (1a) and boundary conditions Eq. (3a) are spatially discretized using a control-volume method on an equidistant $N \times M$ grid. For the explicit integration the second-order Adams-Bashforth time discretization is used in which the first step, for each assimilation interval, is an Euler step. In this explicit implementation (*e4D-Var*) the cost function is first computed by forward evolution over the time interval. The gradient is then evaluated by integrating the adjoint model, with evolution $\mathbf{M}^T(t_i, t_{i-1})$ and forcing $\mathbf{H}_i^T \mathbf{R}_i^{-1} \mathbf{d}_i$, backwards in time. For the implicit integration of the model the second-order Crank-Nicholson scheme is used. When this method is used with 4D-Var, Terwisscha van Scheltinga and Dijkstra (2005) showed that the construction of the adjoint model is straightforward since the transpose of the tangent linear model $\mathbf{M}^T(t_i, t_{i-1})$ is implicitly available from the Newton-Raphson method used to solve the nonlinear systems arising from the Crank-Nicholson method. In the implicit implementation (*i4D-Var*) the cost function in 4D-Var is evaluated by forward evolution over the assimilation interval while simultaneously storing the implicitly available information on \mathbf{M}^T . The gradient is then evaluated by a backward iteration using the stored \mathbf{M}^T . Note that for each backward integration, linear systems have to be solved (Terwisscha van Scheltinga and Dijkstra, 2005); we solve these linear systems with a relative accuracy of $\epsilon_i = 10^{-6}$. Due to the different integration methods in the explicit and implicit schemes, the cost function will also be evaluated differently. This difference depends on the accuracy and stability of the methods involved and therefore depends on the time step used. In all computations below a 60×40 grid was used which was shown to give sufficient accuracy in the solutions computed (Dijkstra and Katsman, 1997). Each linear system (arising from the Newton-Raphson method) in *i4D-Var* was solved using the BiCGSTAB method of Van der Vorst (1989) with an incomplete LU factorization as a preconditioner. For practical implementation of the minimization of the cost function, a NAG-library routine (E04DGF) was used. This method uses a preconditioned conjugate gradient method with a “limited” memory quasi-Newton method to calculate the search direction (Gill and Murray, 1979; Gill et al., 1981). A minimum has been found if the following conditions on the convergence of the iterate, cost function and gradient are *all* satisfied given the optimality

tolerance ϵ_m :

$$J_{k-1} - J_k < \epsilon_m(1 + |J_k|), \quad (13a)$$

$$\|\delta \mathbf{w}_{k-1} - \delta \mathbf{w}_k\| < \epsilon_m^{1/2}(1 + \|\delta \mathbf{w}_k\|), \quad (13b)$$

$$\|\nabla J_k\| \leq \epsilon_m^{1/3}(1 + |J_k|), \quad (13c)$$

where k is the iteration index. For the optimality tolerance in E04DGF, a value of $\epsilon_m=10^{-5}$ was chosen.

3 Comparison between *i*4D-Var and *e*4D-Var

In this section, we will compare the performance of *i*4D-Var and *e*4D-Var using the barotropic quasi-geostrophic ocean model. The specific set-up is described in Sect. 3.1 and results for a standard case, with a fixed time step Δt and a fixed number of points per assimilation interval n , in Sect. 3.2. The changes in performance when Δt and n are varied are presented in Sects. 3.3 and 3.4, respectively and in Sect. 3.5 the overall computational efficiency of 4D-Var in the implicit and explicit model is compared.

3.1 Specific case

For a value of Re far into the irregular regime ($Re=120$), results of a 40 year time integration are shown in Fig. 3a. The quantity $\Delta\Psi$ on the vertical axis in Fig. 3a is a measure of the asymmetry of the streamfunction ψ with respect to the mid axis of the basin and it is defined as:

$$\Delta\Psi = \frac{\max(\psi) + \min(\psi)}{\max(\psi, -\psi)}. \quad (14)$$

A positive value of $\Delta\Psi$ indicates a downward jet-displacement, while a negative value of $\Delta\Psi$ indicates an upward jet-displacement. In these computations, a time step of $\Delta t=15$ min was used for both the implicit and the explicit integration of the model. Both methods give near identical results such that the curves in Fig. 3a are indistinguishable.

The unstable jet-up steady state for $Re=120$ was chosen as the initial state at $t=0$. Although the flow stays close to the initial state for the first few years, the behavior becomes irregular in time with frequent changes between upward and downward jet-displacement. For the comparison between the two 4D-Var implementations, we have derived the “observations” from the 1200 h window after 10 years of integration; the value of $\Delta\Psi$ of these “observations” is plotted in Fig. 3b. Although the computed trajectories were nearly indistinguishable, for consistency the *i*4D-Var observations were taken from the implicit time-integration, while for *e*4D-Var they were taken from the explicit time-integration. The 1200 hour interval of observations is broken into subintervals, each with n points. On each of the subintervals, the minimization problem Eq. (10) is solved with the initial condition as control variables (cf. Fig. 2). For the covariances matrices, we have chosen (for simplicity) that $\mathbf{B}=\mathbf{R}_i=\mathbf{I}$ for

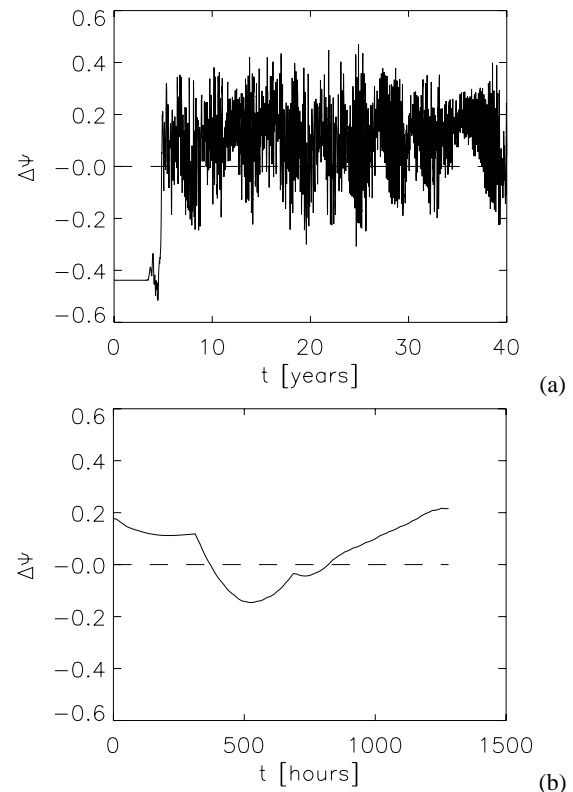


Fig. 3. The asymmetry $\Delta\Psi$ of the streamfunction for: (a) a time-integration of 40 years, starting from a unstable jet-up steady state; and (b) a 1200 h window after year 10 in (a); the latter values serve as the observations.

$i=1, \dots, n$. The identity operator was chosen for the observation operator ($\mathbf{H} = \mathbf{I}$), which means that we use all the observations of the streamfunction. As initial background state we have chosen the unstable jet-up steady state at $Re=120$; this is the starting point of the time-series shown in Fig. 3a. For each interval the first guess of the minimization was taken as $\delta \mathbf{w}=\mathbf{0}$.

3.2 Accuracy

First we consider the behavior of both implementations for a time step $\Delta t=2$ h and $n=2$ (two points per interval). The initial value of the cost function (solid) before minimization and the final value of the cost function after minimization (dashed) are for *e*4D-Var shown in Fig. 4a and for *i*4D-Var in Fig. 4b. Both methods show a rapid decrease in the initial and final value of the cost function for the first few intervals, after which both values becomes constant. The rapid decrease is caused by the large difference at the beginning between the observation (which initially has $\Delta\Psi \geq 0$) and the background state (which initially has $\Delta\Psi \leq 0$). As a result of this difference, both methods make large improvements in the background state until it becomes close to the observa-

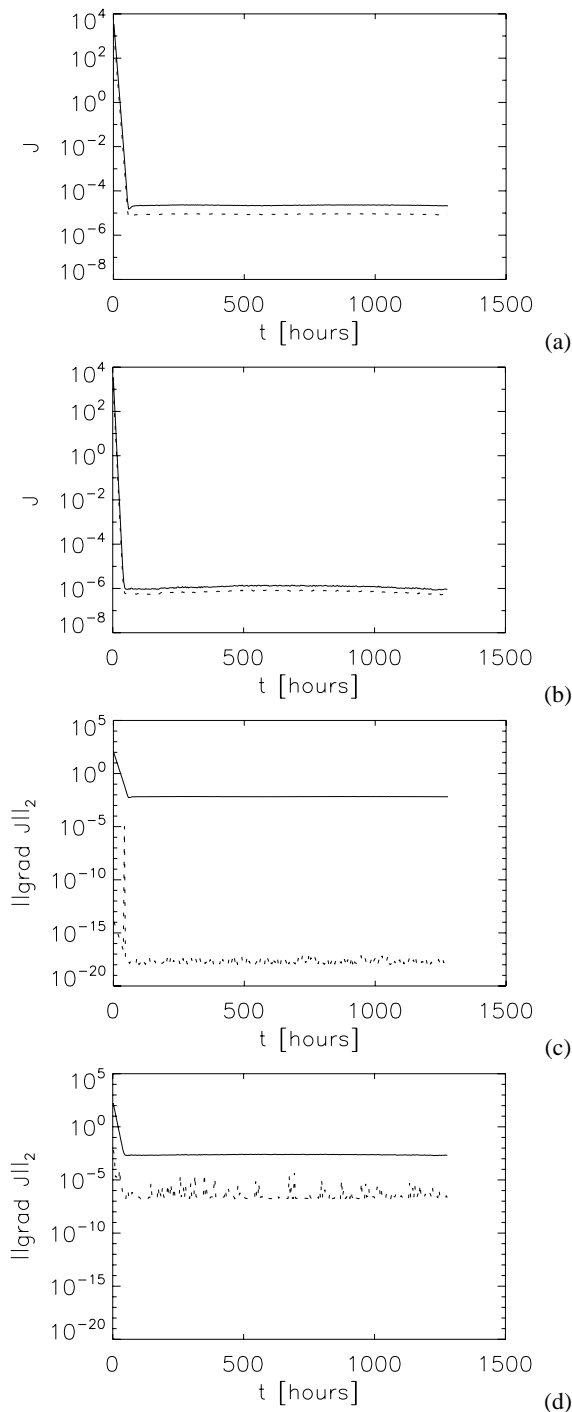


Fig. 4. Results for $\Delta t=2$ h and 2 points per interval. **(a)** Initial value (solid) and final value (dashed) of the cost function for each minimization as evaluated by *e4D-Var*. **(b)** Initial value (solid) and final value (dashed) of the cost function for each minimization as evaluated by *i4D-Var*. **(c)** Initial value (solid) and final value (dashed) of the L_2 norm of the gradient of the cost function for each minimization as evaluated by *e4D-Var*. **(d)** Initial value (solid) and final value (dashed) of the L_2 norm of the gradient of the cost function for each minimization as evaluated by *i4D-Var*.

tions and only a small correction on the background is necessary. The difference in the value of the cost function (both initial and final) between both methods is about one order of magnitude, with the implicit method having the smallest value of the cost function. This is due to the different evaluation of the cost function: for *e4D-Var*, the term involving the tangent linear model in (8) is evaluated at the beginning of each time interval, while in *i4D-Var*, both begin and end points of the interval are used. The initial and final value of the L_2 norm of the gradient ∇J are shown for *e4D-Var* and *i4D-Var* in Fig. 4c and d, respectively. Again there is a sharp decrease initially followed by stabilization afterwards. The initial value of the norm of ∇J is of the same order for both methods but there is a large difference in the magnitude of the final values. This is due to a difference in the evaluation of the gradient: for *i4D-Var*, the evaluation of the gradient requires $2n$ linear systems to be solved which is done using an iterative scheme with an accuracy $\epsilon_i=10^{-6}$. As a result, the L_2 norm of the gradient cannot become smaller than ϵ_i for *i4D-Var*. Since for *e4D-Var* no systems have to be solved, the norm of the gradient can be several orders of magnitude smaller.

An indication of the computational cost for both 4D-Var implementations is provided in Fig. 5a. Here the CPU time (t_{comp}) needed for a minimization over one assimilation interval is plotted for $\Delta t=2$ h and $n=2$ for both *i4D-Var* (solid) and *e4D-Var* (dashed). The *i4D-Var* method is on average a factor 1.5 more expensive in computational time than *e4D-Var*. There are, however, several peaks where the difference is more than a factor 2.5 or higher. For both implementations, the cost function Eq. (7) is minimized using an iterative scheme, with the optimality tolerance $\epsilon_m=10^{-5}$. The conditions on the convergence of the iterate, cost function and gradient are more difficult to satisfy for *i4D-Var*, since the accuracy of the gradient is limited by the tolerance of the iterative linear solver ($\epsilon_i=10^{-6}$). Hence, more iterations are needed for *i4D-Var* than for *e4D-Var* in the optimization procedure. Time integration is also more expensive for *i4D-Var*, since two linear systems have to be solved for each time step: one during evaluation of the cost function and one during evaluation of the gradient. Both factors make *i4D-Var* more expensive than *e4D-Var*.

To summarize the results for the chosen time-step and the number of points per interval: both implementations are capable of finding an accurate analysis. *i4D-Var* appears more accurate than *e4D-Var* for this value of n and Δt , but it is also more expensive.

3.3 Effect of Δt

We use the same setup as in the previous section using the same 1200 h observations but now successively increase the magnitude of the time step, while keeping $n=2$. The initial and final value of the cost function are shown for $\Delta t=2$ h, $\Delta t=4$ h, $\Delta t=8$ h and $\Delta t=16$ h in Fig. 6a–d, respectively.

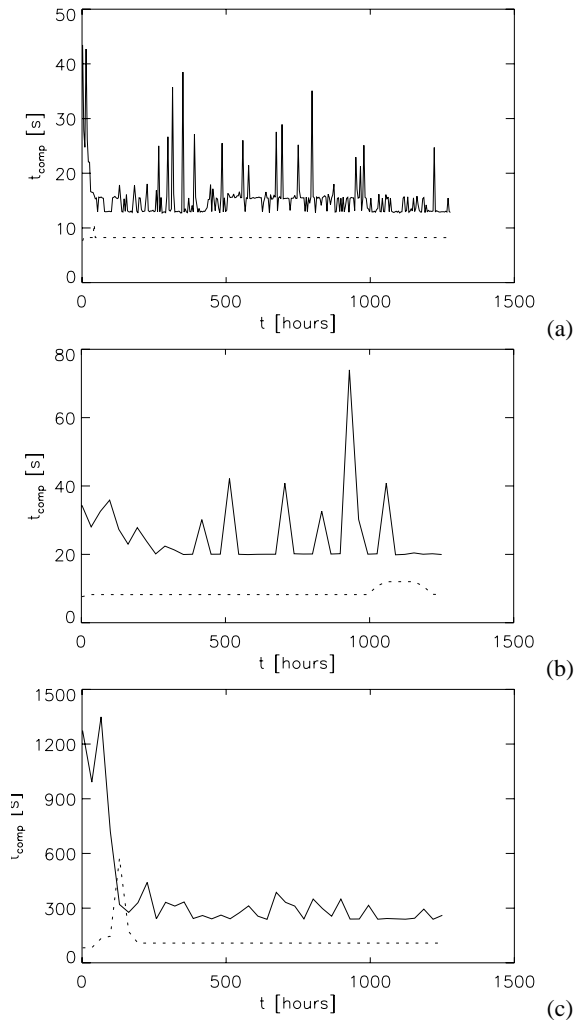


Fig. 5. The processor time t_{comp} needed for the minimization of cost function J for each assimilation interval (see Fig. 2). The value of t_{comp} is plotted at the beginning of the intervals. (a) for $\Delta t = 2$ h and $n = 2$. (b) for $\Delta t = 16$ h and $n = 2$. (c) for $\Delta t = 2$ h and $n = 16$. The solid curves represent the results from $i4D$ -Var. The dashed curves represent the results from $e4D$ -Var.

The latter value of Δt is close to the limiting time step (based on the CFL criterion) of the explicit scheme of $\Delta t \approx 17$ h. In each figure panel, the top two curves are calculated by $e4D$ -Var, while the bottom two are calculated by $i4D$ -Var. For all intervals, the values of the cost function (both initial and final) as calculated by $e4D$ -Var are larger than those calculated by $i4D$ -Var. For the first intervals, the same behavior is observed for both implementations: a rapid decrease of the cost function (both initial and final) and a decrease of the cost function during minimization. After the rapid decrease, the value of the cost function stabilizes.

For $i4D$ -Var, the curves are comparable for each time step, although there is a small increase in the value of both the

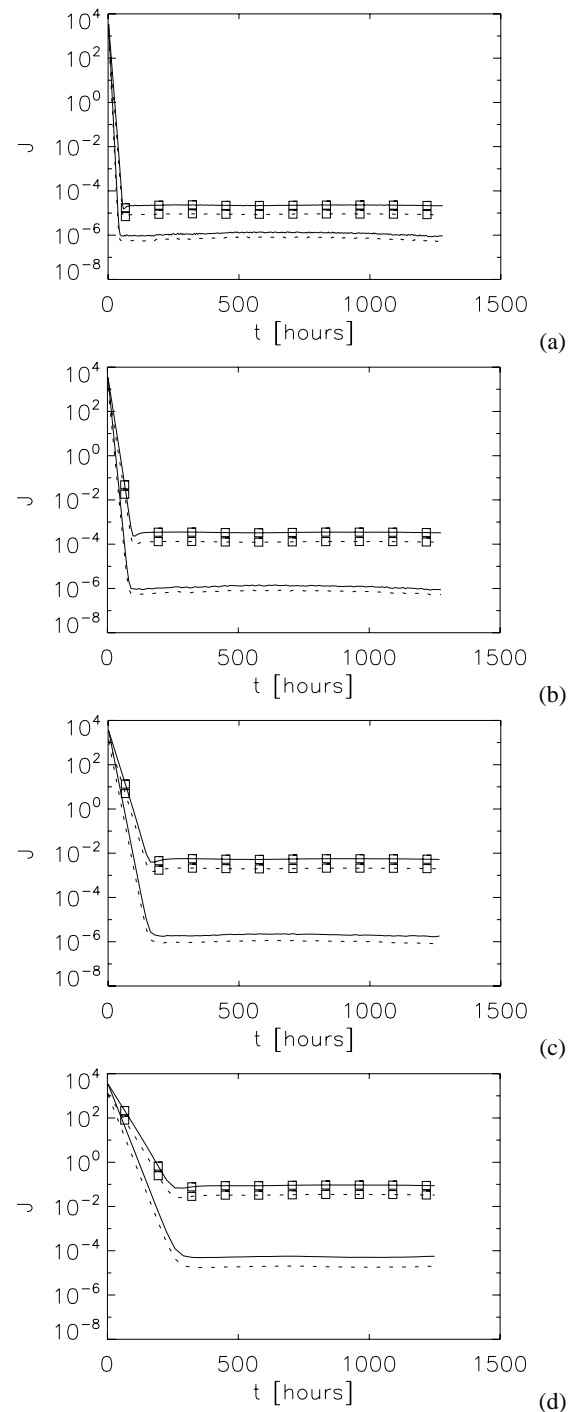


Fig. 6. Value of the initial (solid) and final (dashed) value of the cost function for several values of the time step and $n = 2$. (a) $\Delta t = 2$ h. (b) $\Delta t = 4$ h. (c) $\Delta t = 8$ h. (d) $\Delta t = 16$ h. Curves marked with rectangles denote results of $e4D$ -Var. Curves without rectangles denote results of $i4D$ -Var.

initial and final value of the cost function after the sharp decrease. For $e4D$ -Var, however, the value at which the cost

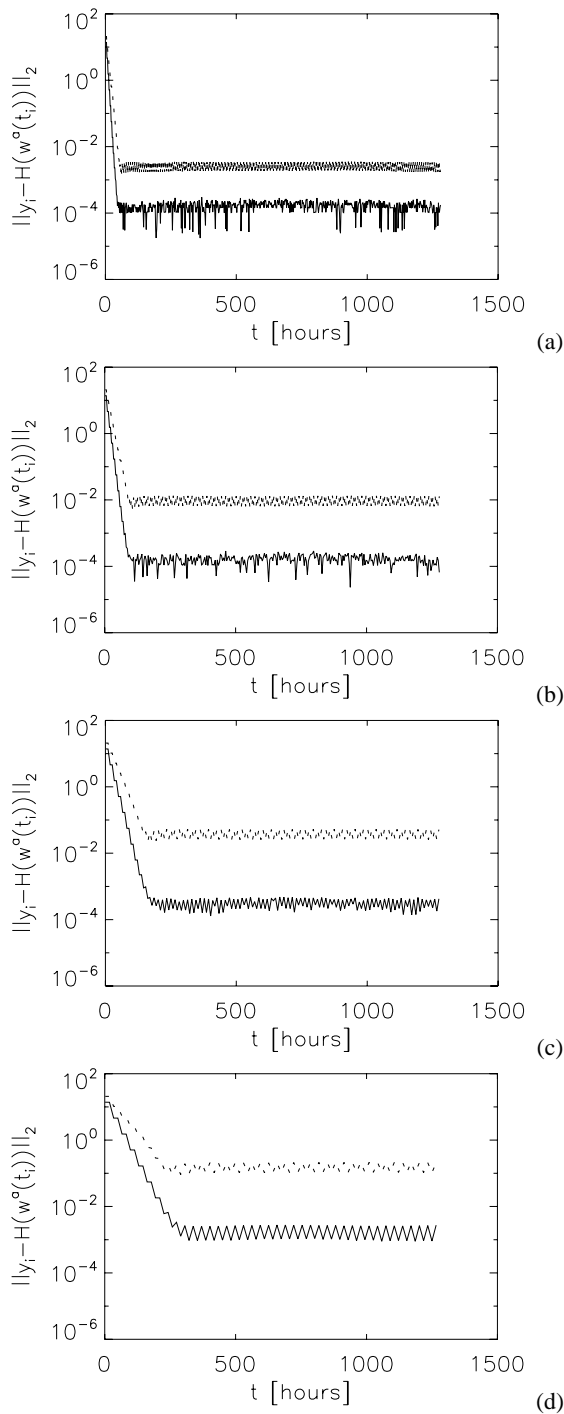


Fig. 7. The difference between the observation and the analysis after minimization of the cost function for several values of the time step. (a) $\Delta t=2$ h. (b) $\Delta t=4$ h. (c) $\Delta t=8$ h. (d) $\Delta t=16$ h. The solid curves represent the results from *i4D-Var*. The dotted curves represent the results from *e4D-Var*.

function stabilizes increases with increasing Δt , for example by 4 orders of magnitude from $\Delta t=2$ h to $\Delta t=16$ h. This is

due to larger error propagation in the explicit scheme used in *e4D-Var* for large Δt . As a result, the evaluation of both the cost function and the gradient becomes less accurate. The same behavior can therefore be seen in the L_2 norm of the difference between analysis and observations, shown in Fig. 7 for different Δt . For both implementations, there is again an increase in the equilibrium value of this norm with Δt but the rate of increase is not as large as for the cost function (Fig. 6). Hence, for increasing Δt the quality of the analysis decreases.

For $\Delta t=16$ h and $n=2$, the CPU time per minimization is shown in Fig. 5b; again *i4D-Var* is more expensive than *e4D-Var*. The difference is on average a factor 2. This is a small increase compared to that found for $\Delta t=2$ h and $n=2$ (Fig. 5a).

3.4 Effect of n

Again using the same set-up as above, we now fix $\Delta t=2$ h and vary the number of points per interval $n=2, 4, 8$ and $n=16$, i.e. the number of observations per subinterval within the 1200 h assimilation interval (cf. Fig. 2). In Fig. 8, the initial and final value of the cost function are plotted for both implementations. Again the two top curves are the results for *e4D-Var*, while the bottom two curves are for *i4D-Var*. In each panel we see a decrease of the cost function in the first few intervals followed by a stabilization. After this decrease there is a difference in behavior: *i4D-Var* is still able to improve the cost function, while *e4D-Var* fails to provide any improvement. The values of the cost function for *e4D-Var* are one order of magnitude larger than those for *i4D-Var* for $n=2$ and this difference increases to six orders of magnitude for $n=16$. For both implementations, the equilibrium values of the cost function increase with n . The rate of increase is one order of magnitude from $n=8$ to $n=16$ for *i4D-Var*, but this is relatively small compared to that of *e4D-Var*.

In Fig. 9, the L_2 norm of the difference between the analysis and the observations is shown for each of value of n (as used in Fig. 8). For *i4D-Var*, this difference decreases in the first interval and then fluctuates around a constant value. With increasing n the results of *i4D-Var* do not change much, apart from a small decrease of the size of the fluctuation and a small increase of the equilibrium value. For *e4D-Var*, however, the equilibrium value does not remain constant with increasing n but it slowly increases with time. In Fig 9b, there is a window in which the L_2 norm strongly fluctuates. To a lesser extent, this is also seen in Fig. 9c but it is absent in Fig. 9d. This window of fluctuations corresponds to a series of observations where the solution changes from jet-down to jet-up and back (Fig. 3b).

The difference between the results of *e4D-Var* and *i4D-Var* can be explained as follows. With increasing n , more integrations for the evaluation of the cost function and its gradient have to be performed. When the number of points per interval is $n+1$, the total number of time steps taken for the eval-

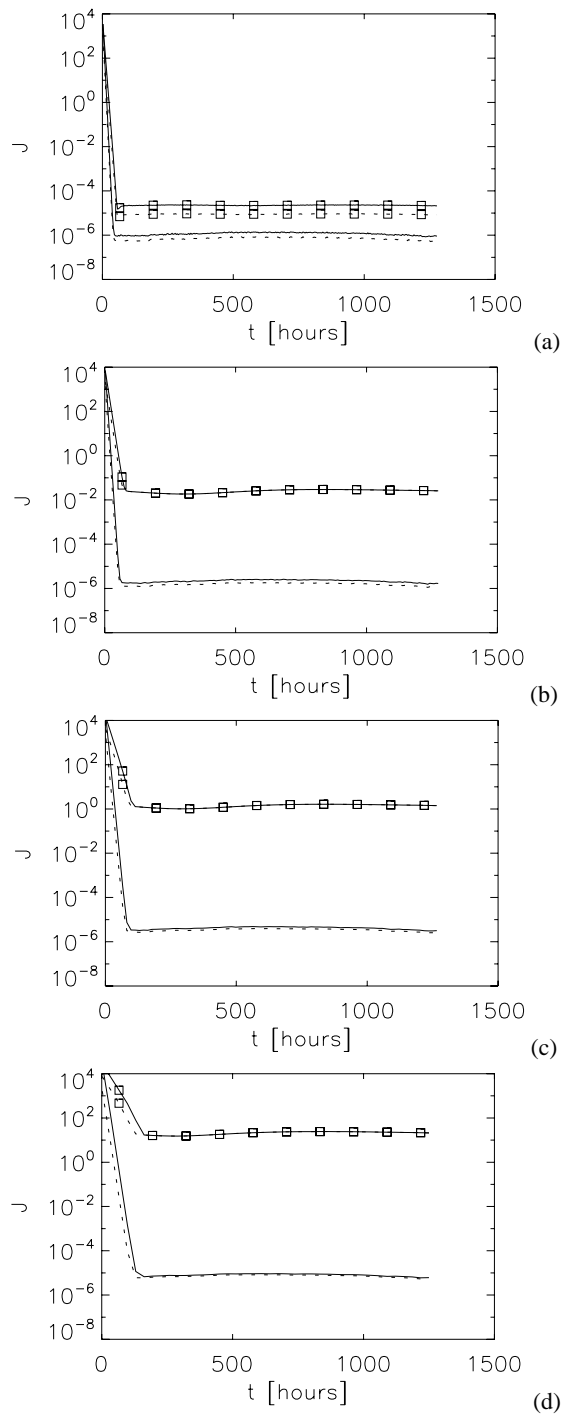


Fig. 8. Value of the initial (solid) and final (dashes) value of the cost function for different n . (a) $n=2$. (b) $n=4$. (c) $n=8$. (d) $n=16$. Curves with rectangles denote results of $e4D$ -Var. Curves without rectangles denote results of $i4D$ -Var.

uation of cost function is n . In $e4D$ -Var, the evaluation of the gradient requires $2n$ steps; n for the forward integration and n for the backward integration with the adjoint model. As

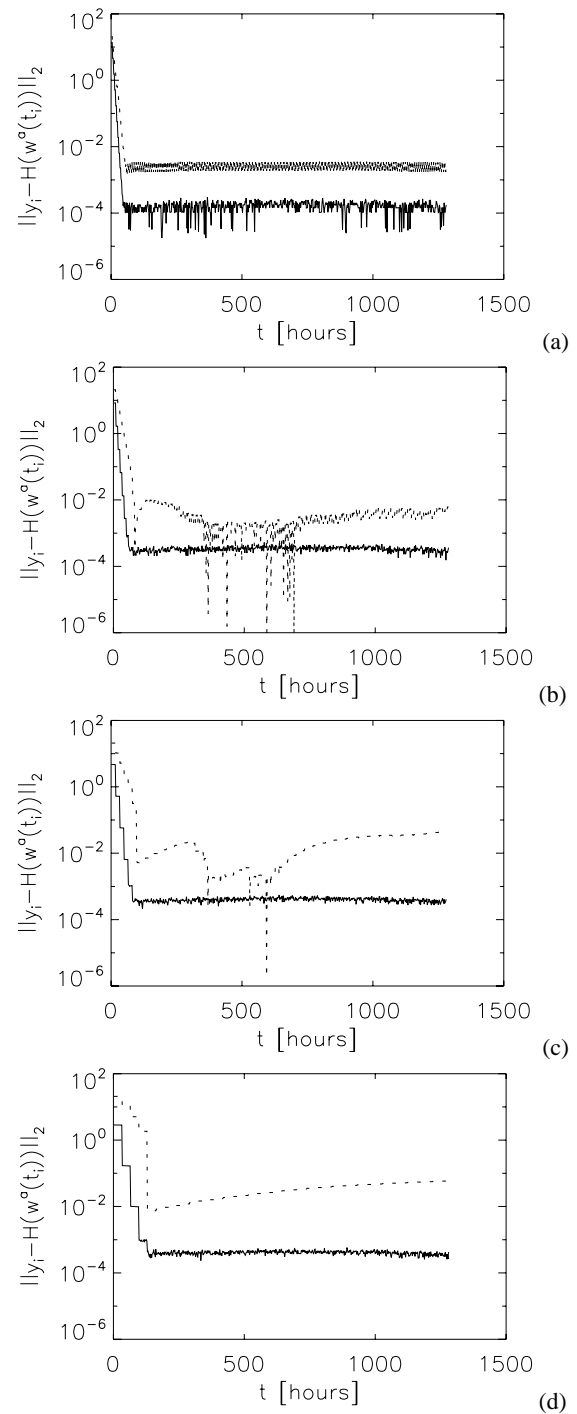


Fig. 9. The difference between the observation and the analysis after minimization of the cost function for different n . (a) $n=2$. (b) $n=4$. (c) $n=8$. (d) $n=16$. The solid curves represent the results from $i4D$ -Var. The dotted curves represent the results from $e4D$ -Var.

a direct result, the cumulative numerical error made in the integration increases. From Fig. 8 it appears that $e4D$ -Var is less accurate than $i4D$ -Var with the same n and the qual-

Table 2. The average accuracy of the analysis $\mathcal{E} = \|\mathbf{y}_i - H(\mathbf{w}^a(t_i))\|_2$ and total processor time Σt_{comp} used for several combinations of the interval length n and time-step Δt and for both explicit and implicit models. For three cases t_{comp} is compared for the implicit and explicit method over the whole interval in Fig. 5; the superscripts a, b and c refer to the subpanels in Fig. 5.

n	Δt [h]	\mathcal{E} (implicit)	Σt_{comp} [s] (implicit)	\mathcal{E} (explicit)	Σt_{comp} [s] (explicit)
2	2	0.0650	4925 ^a	0.1323	2638 ^a
4	2	0.0653	7819	0.2622	7970
8	2	0.0663	10976	0.5190	7299
16	2	0.0782	14570 ^c	0.9819	4868 ^c
2	4	0.1299	3440	0.2684	1347
4	4	0.1306	4553	0.5265	3046
8	4	0.1362	7088	1.0679	4241
16	4	0.1957	9019	2.0794	2918
2	8	0.2598	1678	0.5515	703
4	8	0.2626	2858	1.0695	1045
8	8	0.2989	4909	2.1996	1232
16	8	0.6143	8875	4.0994	1974
2	16	0.5200	1029 ^b	1.1537	349 ^b
4	16	0.5381	1942	2.1787	377
8	16	0.7654	4884	7.3980	548
16	16	2.2575	9237	8.8472	439

ity of the analysis of $e4D$ -Var decreases faster (Fig. 9) with increasing n compared to $i4D$ -Var. For $\Delta t=2$ h and $n=16$, the CPU time per minimization is plotted in Fig. 5c showing that $i4D$ -Var is again more expensive than $e4D$ -Var. After the first few intervals, the minimization scheme terminates after one iteration for $e4D$ -Var since the NAG routine cannot find a direction where the residue is decreased, while conditions on the convergence are not satisfied. The minimization method is unable to find a converged minimum from the initial guess (the initial increment) and the last value provided by the NAG routine is taken as the minimum. This leads to the lack of improvement in the cost function as seen in Figs. 8b–d.

3.5 Overall computational efficiency

In the previous results we saw that $i4D$ -Var was more accurate than $e4D$ -Var but also more expensive. For evaluating whether implicit methods provide a useful alternative for the range of Δt smaller than the maximum value possible with the explicit method, one is interested in a comparison of the total processor time Σt_{comp} needed to obtain a certain average accuracy in the analysis over the whole time interval. As a measure of this average accuracy, we take the quantity \mathcal{E} defined as

$$\mathcal{E} = \overline{\|\mathbf{y}_i - H(\mathbf{w}^a(t_i))\|_2} \quad (15)$$

Here the overbar indicates average of $\|\mathbf{y}_i - H(\mathbf{w}^a(t_i))\|_2$ taken over all the analyses $\mathbf{w}^a(t_i)$ found while assimilating

the observations in the 1200 hour window, using an interval length n and a time-step Δt . The total CPU time Σt_{comp} is the sum of the CPU times needed for each minimization along this interval. In Table 2, values of Σt_{comp} and \mathcal{E} are shown for several combinations of n and Δt and for both explicit and implicit models. The values in Table 2 provide an indication of the computational costs for both methods to produce an analysis with a certain average accuracy. For example with $e4D$ -Var, a value of $\mathcal{E}=0.13$ is achieved for a value of $n=2$ and $\Delta t=2$ h at a computational cost of 2623 seconds. We also see that $i4D$ -Var is more accurate than $e4D$ -Var for the same value of n and Δt but that it is about twice as expensive. To obtain about the same accuracy ($\mathcal{E}=0.13$) with $i4D$ -Var, we can use a larger time step and more points per interval ($\Delta t=2$ and $n=4$) and for this case $i4D$ -Var is only a factor 1.3 (3440/2638) more expensive than $e4D$ -Var.

The values in Table 2 indicate that for $i4D$ -Var, \mathcal{E} does not increase much with n for constant Δt . Only for $\Delta t=16$ there is a large increase for $n=16$, which is due to the relatively large weight of the initial adjustment. For $i4D$ -Var, the total computational time increases approximately linearly with n . For $e4D$ -Var \mathcal{E} always increases with n due to cumulative errors in the time-stepping. For the same Δt , \mathcal{E} for $e4D$ -Var is always larger than that for $i4D$ -Var. The total processor time for $e4D$ -Var varies non-monotonically with increasing n . This is because for large n and Δt the minimization terminates unsuccessfully due to inaccuracies in the integration method. From Table 2, we also see that for the particular model used here, $i4D$ -Var can be more efficient than $e4D$ -Var even in the range of values of Δt below the CFL limit. For example, if a value of $\mathcal{E}=0.52$ is desired, we could use $n=2$ and $\Delta t=16$ h for $i4D$ -Var which would cost 1029 s. For the same n , we would have to use a $\Delta t=4$ h with $e4D$ -Var to obtain approximately the same value of \mathcal{E} , which would cost 1347 s.

4 Performance of $i4D$ -Var

In this section, we will investigate the performance of $i4D$ -Var for a time step $\Delta t=24$ h, which is larger than the maximum time step possible for the Adams-Bashforth scheme and hence (for this time step) $e4D$ -var does not work. As the target problems for $i4D$ -Var are those where variability is on relatively long time scales or equilibrium behavior occurs, we consider flow regimes in the barotropic ocean model where steady states exist. We use $n=5$ in all computations below and will investigate two cases: (i) a model mismatch, (ii) a solution mismatch and (iii) the effect of observational noise. The aim of (i) is to investigate whether the method finds an analysis close to the observations, using a ‘badly tuned’ model. We consider the regime $Re < 30$ where we know that there is a unique solution for which the barotropic streamfunction is anti-symmetric (cf. Fig. 1a). As ‘observations’, we take a daily sequence of the steady-state solution

of the model for $Re=25$. The background model is initialized with the steady-state solution at $Re=20$ and also run for $Re=20$. With $i4D$ -Var approach, we assimilate observations obtained at one value of Re within the model which is run with a “wrong” value for Re . The initial increments for each assimilation interval are taken equal to zero and the observation error covariance matrices \mathbf{R}_i , $i=1, \dots, n$ are taken equal to the identity matrix \mathbf{I} . A large decrease between the initial and final values of the cost function (Fig. 10a) and the L_2 norm of the gradient (Fig. 10b) occurs. After the first four intervals, the value of the cost function stabilizes and the assimilation method cannot improve the analysis anymore. The difference between the observations and the analysis also stabilizes after a few intervals (Fig. 10c). This stabilization at a relatively large error is due to the fact that the background model has a unique steady solution at $Re=20$. Therefore, it is not possible to find an analysis that perfectly fits the observations (for $Re=25$), since the observations are derived from a different steady solution. Although the ‘wrong’ background model will not allow a perfect fit, $i4D$ -Var finds an analysis which improves the solution of the “badly tuned” model to be closer to the observations.

Next, we consider how 4D-Var performs under case (ii) of a solution mismatch. As discussed earlier, for $30 < Re < 52$, the barotropic ocean model has two stable steady solutions (cf. Fig. 1b, c). For $Re=50$, which is in this multiple equilibria regime, we synthesize “observations” from the jet-up solution in Fig. 1b, while we initialize the background model with the jet-down solution in Fig. 1c. The aim is to test whether the assimilation method is able to find the correct stable equilibrium, while being initialized with the “wrong one”. The initial increment for each interval is again taken equal to be zero and again the covariance matrices \mathbf{R}_i , $i=1, \dots, n$ are taken to be the identity matrix \mathbf{I} . For each interval, there is a large difference between the initial and final value of the cost function (Fig. 11a) and the L_2 norm of its gradient (Fig. 11b). A decrease of the initial and final values for each successive interval occurs, indicating that the analysis converges towards the observations. In Fig. 11c, the norm of the difference between the observations and the analysis converges towards zero. This indicates that $i4D$ -Var is able to find an analysis which is a perfect fit to the observations in this case.

Finally, we consider case (iii) for which noise is added to the observations. The steady state ψ for $Re=25$ is perturbed by adding noise to obtain the “observations”

$$y_i = \psi + N_i(\mathbf{0}, \max |\psi| \mathbf{I}), \quad (16)$$

where the maximum is taken over all the gridpoints and each $N_i(\mu, \mathbf{C})$ is a Gaussian distribution, where μ is the mean and \mathbf{C} the covariance matrix. The model is run (with a time-step of 24 h) at $Re=25$ and initialized with the steady-state solution ψ . For every assimilation interval, the initial increment is taken equal to zero and the observation error covari-

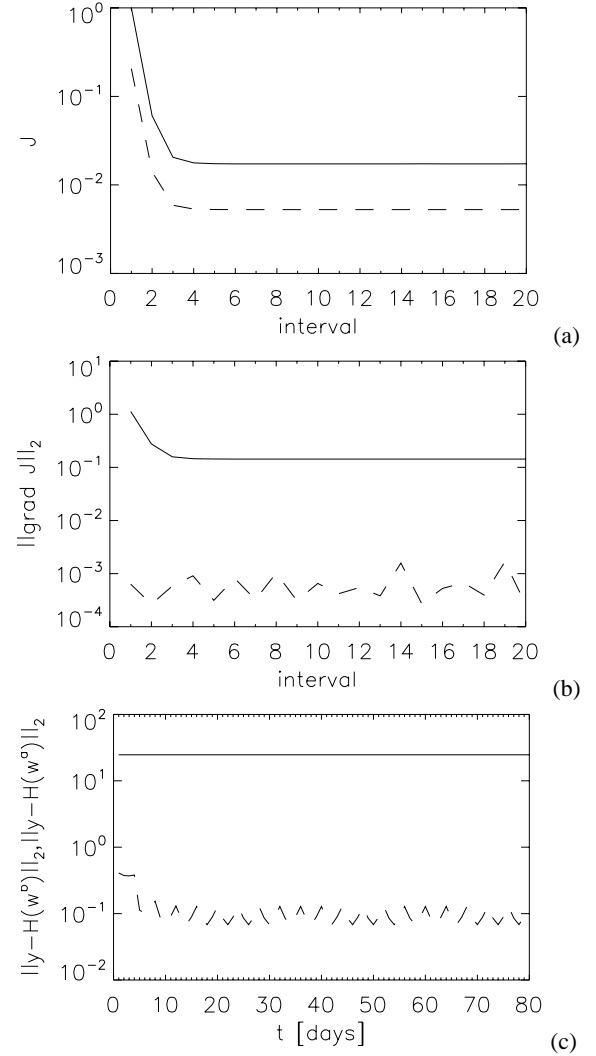


Fig. 10. The solution mismatch case for $Re=50$. In this case, the “observations” are from Fig. 1c while the model is initialized with the solution in Fig. 1b. **(a)** Initial value (solid) and final value (dashed) of the cost function. **(b)** Initial value (solid) and final value (dashed) of the norm of the gradient. **(c)** Norms of the difference between the data and the model (solid) without assimilation and the data and the analysis (dashed) after assimilation.

ance matrix is taken as the covariance of the noise, i.e. for $i=1, \dots, n$,

$$\mathbf{R}_i = \max |\psi| \mathbf{I}. \quad (17)$$

The results in Fig. 12 indicate that there is a large difference between the initial and the final values of the cost function (Fig. 12a) and the norm of the gradient (Fig. 12b). The reason for the increase of the initial value of the cost function after the first interval is that the analysis found for the each assimilation interval is not a solution of the model equations. The background term $H_i M(t_i, t_0)(\mathbf{w}^b(t_0))$ in (8) therefore

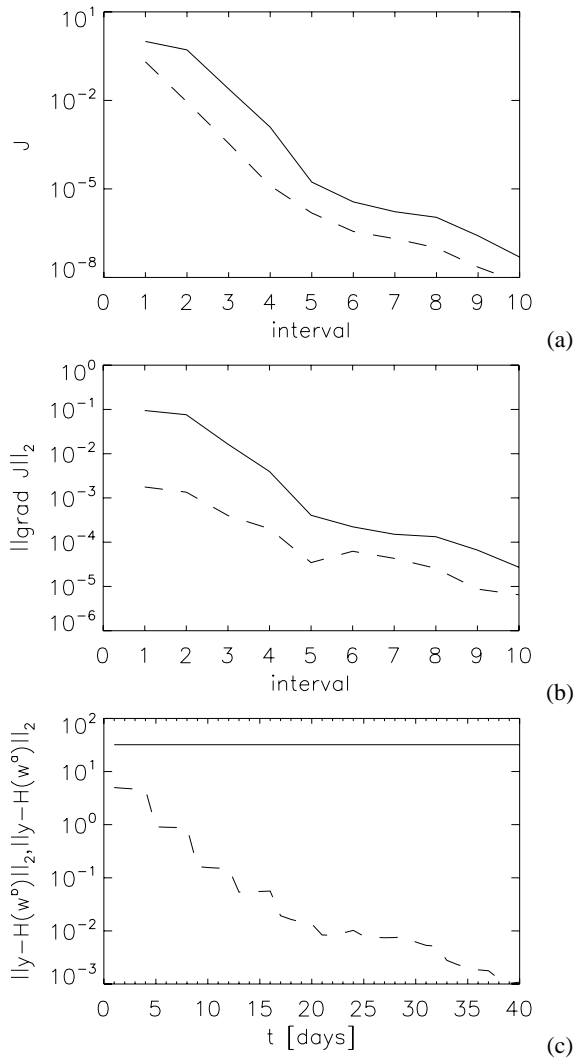


Fig. 11. The solution mismatch case for $Re=50$. In this case, the ‘observations’ are from Fig. 1c while the model is initialized with the solution in Fig. 1b. **(a)** Initial value (solid) and final value (dashed) of the cost function. **(b)** Initial value (solid) and final value (dashed) of the norm of the gradient. **(c)** Norms of the difference between the data and the model (solid) without assimilation and the data and the analysis (dashed) after assimilation.

increases after the first interval. From the norm of the difference between the observations and the model solutions before and after assimilation (Fig. 12c), it is seen that the analysis is much closer to the observations than the model solutions were before assimilation. The analysis at the beginning of the interval differs from the background state. Since the background model is used as a strong constraint, it will pull the analysis at the other points from the observations towards the background state. This influence becomes stronger towards the end of the interval and will result in an increase in the difference between the observations and the analysis (Fig. 12c).

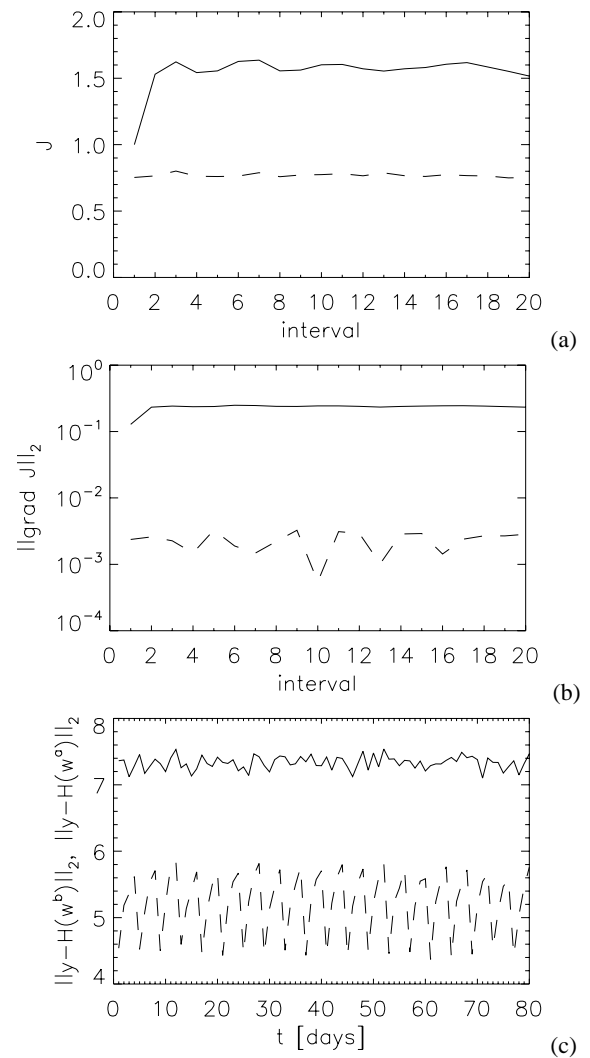


Fig. 12. Identical-twin experiment with Gaussian noise for $Re=25$. **(a)** Initial value (solid) and final value (dashed) of the cost function. **(b)** Initial value (solid) and final value (dashed) of the norm of the gradient. **(c)** Norms of the difference between the data and the model (solid) without assimilation and the data and the analysis (dashed) after assimilation.

5 Conclusions

Results were presented of a comparison between the performance of 4D-Var in an explicit (Adams-Bashforth) and an implicit (Crank-Nicholson) version of a nonlinear barotropic quasi-geostrophic ocean model for which ‘observations’ were derived from solutions of this model. Several flow regimes of the model, depending on the Reynolds number Re , were considered. At high values of Re , the flow is highly irregular and is associated with rapid transitions between jet-up and jet-down flows. In this irregular flow regime, both 4D-Var implementations are capable of producing accurate

analyses given the observations. Increasing the size of the time step Δt , or the number of points per interval n , leads to reduced quality of the analysis for *e*4D-Var when compared to *i*4D-Var. This result is due to cumulative numerical inaccuracies which occur in the explicit time-stepping scheme. Apart from that, even with the same Δt and n , the analyses from *i*4D-Var are more accurate due to the more accurate evaluation of the cost function. It was demonstrated (cf. Table 2) that *i*4D-Var can be a more efficient method (smaller total computational time) than *e*4D-Var since it is more accurate at larger time steps. At smaller values of Re stable steady states exist in the model and the flows display near-equilibrium behavior. In this regime, *i*4D-Var offers the possibility to use time steps much larger than those possible for *e*4D-Var. We showed that *i*4D-Var is capable of finding an accurate analysis when the background model is “badly tuned”. Furthermore, in a regime of multiple steady-state solutions, *i*4D-Var is capable of finding an analysis which is a near perfect fit to the “correct” equilibrium, when initialized with the “wrong” one. Also for “noisy” observations, *i*4D-Var performs well and Terwisscha van Scheltinga and Dijkstra (2005) found that *i*4D-Var is also capable of accurately estimating multiple parameter values. Together with the fact that no explicit adjoint has to be constructed, *i*4D-Var can be an attractive alternative for data-assimilation problems in which flows are changing much more slowly than on a time scale comparable to the maximum time step allowed by numerical stability in explicit models. We admit, however, that the development of implicit methods for large scale ocean models is still in its infancy. The model problem chosen here, with the idealized observations and identity observation and covariance matrices, is orders of magnitude simpler than most ocean models used in operational oceanography. Two concerns to the operational applicability of the *i*4D-Var approach may come to mind: (i) the implicit model may be as difficult to construct as the adjoint model, and (ii) the computational costs and storage requirements become prohibitively expensive for systems with a larger number of degrees of freedom. A discussion of both concerns for further development of *i*4D-Var towards its application to real world situations is therefore warranted. Concerning (i), our experience with the development of 3D primitive equation ocean models is that the tangent linear model (or Jacobian matrix) of the set of nonlinear equations arising from an implicit discretization technique can be computed for complicated ocean model formulations. Our approach is to determine the tangent linear model first locally using the stencil defined by the spatial discretization and then building up the total matrix elementwise as in a finite element method. When the local coupling between different unknowns becomes complicated, such as the implementation of neutral physics, that part of the local Jacobian is computed using numerical differentiation. Having developed a 3D global ocean model this way (Weijer et al., 2003), we think that the Jacobian matrix can be determined for ocean models which

are now used in operational oceanography. With this Jacobian matrix, the gradient of the cost function can be determined in *i*4D-Var with the use of in situ transposition (Saad, 1994) of the tangent linear model (as is used here also for the barotropic quasi-geostrophic model). The issue (ii) is more complicated. In the Crank-Nicholson method (or any other implicit time stepping scheme), nonlinear systems of equations have to be solved with the Newton-Raphson method (or any quasi-Newton method, such as the adaptive Shamanskii method (Weijer et al., 2003)). To do this, efficient linear system solvers are required. For the barotropic quasi-geostrophic model as used here, such a solver is easily available but for more complicated ocean models at higher resolution, the development of these solvers is a complicated problem (Dijkstra, 2005). There has been much progress, however, to develop targeted solvers for primitive equation ocean models. The recently developed block Gauss-Seidel preconditioner (De Niet et al., 2007) allows to efficiently solve systems of equations having up to 2 million degrees of freedom with the GMRES technique (Saad, 1994). These solvers will also increase the application potential of 4D-Var to implicit models.

Acknowledgements. This work was supported by the Dutch Technology Foundation (STW) within the project GWI.5798.

Edited by: J. Kurt

Reviewed by: one anonymous referee

References

- Courtier, P., Thépaut, F.-N., and Hollingsworth, A.: A strategy for operational implementation of 4D-Var, using an incremental approach, *Q. J. Roy. Meteor. Soc.*, 120, 1367–1388, 1994.
- De Niet, A., Wubs, F., Dijkstra, H., and Terwisscha van Scheltinga, A.: A tailored solver for bifurcation analysis of high-resolution ocean models, *J. Comp. Phys.*, 227, 654–679, 2007.
- Dijkstra, H. A.: *Nonlinear Physical Oceanography: A Dynamical Systems Approach to the Large Scale Ocean Circulation and El Niño*, Springer, Dordrecht, The Netherlands, 2nd edn., 532 pp., 2005.
- Dijkstra, H. A. and Katsman, C. A.: Temporal variability of the Wind-Driven Quasi-geostrophic Double Gyre Ocean Circulation: Basic Bifurcation Diagrams, *Geophys. Astrophys. Fluid Dyn.*, 85, 195–232, 1997.
- Dijkstra, H. A., te Raa, L., and Weijer, W.: A systematic approach to determine thresholds of the ocean’s thermohaline circulation, *Tellus A*, 56, 362–370, 2004.
- Ferreira, D., Marshall, J., and Heimbach, P.: Estimating Eddy Stresses by Fitting Dynamics to Observations Using a Residual-Mean Ocean Circulation Model and Its Adjoint, *J. Phys. Ocean.*, 35, 1891–1910, 2005.
- Giering, R. and Kaminski, T.: Recipes for adjoint code construction, *ACM Trans. Math. Softw.*, 24, 437–474, 1998.
- Gill, P. and Murray, W.: The computation of Lagrange multiplier estimates for constrained minimization, *Math. Program.*, 17, 32–60, 1979.

- Gill, P., Wright, M., and Murray, W.: Practical Optimization, Academy Press, 401 pp., 1981.
- Ide, K., Courtier, P., Ghil, M., and Lorenc, A.: Unified notation for data assimilation: operational, sequential and variational, *J. Meteor. Soc. Japan*, 75, 181–189, 1997.
- Klinker, E., Rabier, F., Kelly, G., and Mahfouf, J.-F.: The ECMWF operational implementation of four dimensional variational assimilation. Part III: Experimental results and diagnostics with operational configuration, *Q. J. Roy. Meteor. Soc.*, 126, 1191–1215, 2000.
- Mahfouf, J.-F. and Rabier, F.: The ECMWF operational implementation of four dimensional variational assimilation. Part II: Experimental results with improved physics, *Q. J. Roy. Meteor. Soc.*, 126, 1171–1190, 2000.
- Pedlosky, J.: Geophysical Fluid Dynamics, 2nd Edn., Springer-Verlag, New York, 710 pp., 1987.
- Rabier, F., Järvinen, H., Klinker, E., Mahfouf, J.-F., and Simmons, A.: The ECMWF operational implementation of four dimensional variational assimilation. Part I: Experimental results with simplified physics, *Q. J. Roy. Meteor. Soc.*, 126, 1143–1170, 2000.
- Saad, Y.: SPARSKIT: A basic tool kit for sparse matrix computations, Tech. Rep., Computer Science Department, University of Minnesota, 27 pp., 1994.
- Schmeits, M. J. and Dijkstra, H. A.: On the physics of the 9 months variability in the Gulf Stream region: combining data and dynamical systems analysis, *J. Phys. Oceanogr.*, 30, 1967–1987, 2000.
- Simonnet, E., Ghil, M., and Dijkstra, H.: Homoclinic bifurcations in the quasi-geostrophic double-gyre circulation, *J. Mar. Res.*, 63, 931–956, 2005.
- Stammer, D.: Adjusting internal model errors through ocean state estimation, *J. Phys. Oceanogr.*, 108, 1143–1153, 2005.
- Stammer, D., Davis, R., Fu, L.-L., Fukumori, I., Giering, R., Marotzke, J., Marshall, J., Menemenlis, D., Niiler, P., Wunsch, C., and Zlotnicki, V.: Ocean state estimation in support of CLIVAR and GODAE, *CLIVAR Exchanges*, 5, 3–5, 2000.
- Stammer, D., Wunsch, C., Fukumori, I., and Marshall, J.: State estimation in modern oceanographic research, *Eos, Trans. Amer. Geophys. Union*, 83, 289, 294–295, 2002a.
- Stammer, D., Wunsch, C., Giering, R., Eckert, C., Heimbach, P., Marotzke, J., Adcroft, A., Hill, C., and Marshall, J.: The global ocean circulation 1992–1997, estimated from a general circulation model constrained by WOCE data, *J. Geophys. Res.*, 107, 3007, doi:10.1029/2001JC000888, 2002b.
- Stammer, D., Wunsch, C., Giering, R., Eckert, C., Heimbach, P., Marotzke, J., Adcroft, A., Hill, C., and Marshall, J.: Volume, heat and freshwater transports of the global ocean circulation 1993–2000, estimated from a general circulation model constrained by WOCE data, *J. Geophys. Res.*, 108, C05023, doi:10.1029/2001JC001115, 2003.
- Terwisscha van Scheltinga, A. and Dijkstra, H.: Nonlinear data-assimilation using implicit models, *Nonlin. Processes Geophys.*, 12, 515–525, 2005, <http://www.nonlin-processes-geophys.net/12/515/2005/>.
- Van der Vorst, H.: Bi-CGSTAB, a fast and smoothly converging variant of Bi-CG for the solution of nonsymmetric linear systems, *SIAM J. Sci. Statist. Comput.*, 13, 631–644, 1989.
- Vialard, J., Weaver, A., Anderson, D., and Delecluse, P.: Three- and four-dimensional variational assimilation with an ocean general circulation model of the tropical Pacific. Part II: physical validation, *Mon. Weather Rev.*, 131, 1379–1395, 2003.
- Weaver, A., Vialard, J., Anderson, D., and Delecluse, P.: Three- and four-dimensional variational assimilation with an ocean general circulation model of the tropical Pacific. Part I: formulation, internal diagnostics and consistency checks, *Mon. Weather Rev.*, 131, 1360–1378, 2003.
- Weijer, W., Dijkstra, H. A., Oksuzoglu, H., Wubs, F., and De Niet, A.: A fully-implicit model of the global ocean circulation, *J. Comp. Phys.*, 192, 452–470, 2003.
- Wenzel, M., Schröter, J., and Olbers, D.: The annual cycle of the global ocean circulation as determined by 4D-Var data assimilation, *Prog. Ocean.*, 48, 73–119, 2001.
- Winguth, A., Archer, D., Maier-Reimer, E., and Mikolajewicz, U.: Paleonutrient data analysis of the glacial Atlantic using adjoint ocean general circulation models, in: *Inverse methods in global biogeochemical cycles*, edited by: Kasibhatla, P., Heimann, M., Harley, D., Mahowald, N., Prinn, R., and Rainer, P., *Geophysical Monograph Series*, American Geophysical Union, 171–183, 2000.